A Model of Bias in Lake Selection for Survey

C.K. Minns

Great Lakes Fisheries Research Branch,
Canada Centre for Inland Waters,
867 Lakeshore Road, P.O. Box 5050
Burlington, Ontario. L7R 4A6

December 1986

Canadian Technical Report of Fisheries and Aquatic Sciences

No. 1496

87-02245/

Canadian Technical Report of Hydrography and Ocean Sciences

Technical reports contain scientific and technical information that contributes to existing knowledge but which is not normally appropriate for primary literature. The subject matter is related generally to programs and interests of the Ocean Science and Surveys (OSS) sector of the Department of Fisheries and Oceans.

Technical reports may be cited as full publications. The correct citation appears above the abstract of each report. Each report is abstracted in *Aquatic Sciences and Fisheries Abstracts* and indexed in the Department's annual index to scientific and technical publications.

Technical reports are produced regionally but are numbered nationally. Requests for individual reports will be filled by the issuing establishment listed on the front cover and title page. Out of stock reports will be supplied for a fee by commercial agents.

Regional and headquarters establishments of Ocean Science and Surveys ceased publication of their various report series as of December 1981. A complete listing of these publications is published in the *Canadian Journal of Fisheries and Aquatic Sciences*, Volume 39: Index to Publications 1982. The current series, which begins with report number 1, was initiated in January 1982.

Rapport technique canadien sur l'hydrographie et les sciences océaniques

Les rapports techniques contiennent des renseignements scientifiques et techniques qui constituent une contribution aux connaissances actuelles, mais qui ne sont pas normalement appropriés pour la publication dans un journal scientifique. Le sujet est généralement lié aux programmes et intérêts du service des Sciences et levés océaniques (SLO) du ministère des Pêches et des Océans.

Les rapports techniques peuvent être cités comme des publications complètes. Le titre exact paraît au-dessus du résumé de chaque rapport. Les rapports techniques sont résumés dans la revue *Résumés des sciences aquatiques et halieutiques*, et ils sont classés dans l'index annuel des publications scientifiques et techniques du Ministère.

Les rapports techniques sont produits à l'échelon régional, mais numérotés à l'échelon national. Les demandes de rapports seront satisfaites par l'établissement auteur dont le nom figure sur la couverture et la page du titre. Les rapports épuisés seront fournis contre rétribution par des agents commerciaux.

Les établissements des Sciences et levés océaniques dans les régions et à l'administration centrale ont cessé de publier leurs diverses séries de rapports en décembre 1981. Une liste complète de ces publications figure dans le volume 39, Index des publications 1982 du *Journal canadien des sciences halieutiques et aquatiques*. La série actuelle a commencé avec la publication du rapport numéro 1 en janvier 1982.

CANADIAN TECHNICAL REPORT OF FISHERIES

AND AQUATIC SCIENCES 1496

November 1986

A MODEL OF BIAS IN LAKE SELECTION FOR SURVEY

by

C. K. Minns

Great Lakes Fisheries Research Branch

Canada Centre for Inland Waters

867 Lakeshore Road, P.O. Box 5050

Burlington, Ontario L7R 4A6

(c) Minister of Supply and Services Canada, 1986
Cat. No. FS 97-6/1496 ISSN 0706-6457

Correct citation for this publication:

Minns, C.K. 1986. A model of bias in lake selection for survey. Can. Tech. Rep. Fish. Aquat. Sci., 1496:i-v, 1-21pp.

ABSTRACT

Minns, C. K. 1986. A model of bias in lake selection for survey. Can. Tech. Rep. Fish. Aquat. Sci., 1496:i-v, 1-21pp.

The purpose of this report is to describe the results of analyses of the Ontario lake inventory data base. Models of selection bias are developed from a model of predator preference. The models are used to estimate the selection bias toward lakes containing four gamefish species, brook trout, lake trout, smallmouth bass and walleye. The results suggest that co-occurrences of these species are overrepresented in the data base and that unsurveyed lakes, if they have gamefish, will likely only have one of the four species present. Alternative schemes for eliminating or accommodating the biases are discussed.

RESUME

Minns, C. K. 1986. A model of bias in lake selection for survey. Can. Tech. Rep. Fish. Aquat. Sci., 1496:i-v, 1-21pp.

Le présent rapport vise à décrire les résultats des analyses de la base de données de l'inventaire des espèces de poissons du lac Ontario. Des modèles d'erreur systématique de choix sont élaborés à partir d'un modèle de préférence des prédateurs. Les modèles servent à évaluer l'erreur systématique de choix à l'égard de lacs contenant quatre espèces de poissons-gibier (omble de fontaine, touladi, achigan à petite bouche et doré). D'après les résultats, la présence simultanée de ces espèces est sur-représentée dans la base de données et il est probable qu'advenant la présence de poissons-gibier dans les lacs non étudiés, seule une des quatre espèces sera présente. D'autres plans pour éliminer ou adapter les erreurs sont abordés.

LIST OF TABLES

Table 1 Estimates of the bias coefficient α given n_a, n_t*, N_t* and 'guestimates' of N_a for four species of gamefish, using equation 4.

Table 2 Expected sample of lakes containing species a and b (n_{ab}) after biased sampling of 500 (n_t) of 10,000 (N_t) lakes for various combinations of α and β using equations 7-10. (The line delimits those combinations where all available lakes with a and b are sampled.)

Table 3 Percentage of sampled lakes with both species present versus cumulative sample size with Na=100, Nb=100, Nab=10, Nt=10,000 using equations 7-10.

Table 4 Results of linear regressions of na/nt vs nt/Nt for lake trout by primary watershed and lake size range in Ontario, with and without zero values (na/nt=0).

Table 5 Observed cumulative number of lake inventories, percentage of lakes with one or more gamefish species and percentage of lakes with one of fifteen combinations of gamefish presence/absence.

Table 6 Assumed occurrences (Ni) of all combinations of the four gamefish species in Ontario lakes.

Table 7 Predicted percentage of lakes with one or more gamefish species and percentage of lakes with one of fifteen combinations of gamefish presence/absence, using the biases in Table 1.

INTRODUCTION

According to Cox (1978) the province of Ontario has more than 180,000 lakes, each with an area of one ha. or greater. While there are a number of large lakes including the Great Lakes, the vast majority (94.1 percent) have an area less than 100 ha. Of all these lakes, approximately 9,000 have been inventoried by the Ontario Ministry of Natural Resources. Inventories are conducted using a standard procedure (Dodge et al. 1981) and are designed basically to provide morphometric and physico-chemical data plus a fish species list.

The collection of inventories conducted to date is not a random subsample of total population of lakes (Goodchild and Gale 1981). The database is biased toward large lakes (Minns 1984) and toward lakes containing certain preferred gamefish species (Minns 1981). Those species include brook trout, Salvelinus fontinalis, lake trout, Salvelinus namaycush, smallmouth bass, Micropterus dolomieui, and walleye, Stizostedion vitreum vitreum.

In recent years, the database has come to be recognized as an important source of information for a number of resource management issues (Goodchild and Gale, 1981). Those issues include: selection of lakes as class representatives for a province-wide assessment network (OMNR 1978); and identification of the potential extensive impact of acidic deposition on valued fishery resources (Minns 1981).

To extrapolate from the inventory sample to the total lake set, the database must be assumed to be representative of the total set. For instance, Minns (1981) found that 1496 out of 6393 inventoried lakes contained brook trout. If this was assumed to be representative of the total population, one

would estimate there were 42359 brook trout lakes in Ontario. However, an exhaustive search for brook trout lakes using informal information available in district and regional offices of OMNR, museum records, etc., produced an estimated total of 2110 brook trout lakes (C.H. Olver quoted in Minns 1981). Similar results were indicated for other gamefish species (Minns 1981). As the database inventory is a biased sample of the total lake set, it is essential that quantified means of correcting for biases be developed before further extrapolations are attempted.

The purposes of this report are: (1) to examine simple models of biased sampling; (2) to present the results of exploratory attempts to quantify bias in the inventory database; and (3) to suggest future sampling options whereby the bias problem may be made more manageable.

BIASED SAMPLING MODELS

Lakes selected for inventory are rarely resurveyed. This represents a situation of selection from a mixed population without replacement. Manly et al. (1972) developed a model to describe this phenomenon which allows for preferences, or biased selection. Recast in terms of biased selection of lakes for inventory, the model is as follows: suppose there is a total set of N_t lakes of which N_a contain fish species a. A lake containing species a is (relative preference) times more likely to be selected for inventory. If the numbers of lakes with and without species a are equal, the probability of picking a lake with a is

$$P(a) = \frac{\alpha}{1 + \alpha} \qquad (1)$$

If the numbers of lakes are unequal then the coefficients must be weighted by their corresponding population sizes. Thus,

$$P(a) = \frac{Na\alpha}{(Nt-Na) + Na\alpha} \qquad (2)$$

for the first lake selected. On successive selections, since lakes are not inventoried a second time, the probabilities change. After n_t lakes have been inventoried, of which n_a contained species a, the probability that the next lake inventoried contains a is

$$P(a) = \frac{(Na-na)\,\alpha}{[(Nt-Na)-(nt-na)] + (Na-na)\alpha} \qquad (3)$$

Manly et al (1972) showed that this equation could not be integrated with respect to n_t. However, by means of approximation, they showed that an approximation could be developed as a basis for estimating α from known data:

$$1 - \frac{nt-na}{Nt-Na} = \left(1 - \frac{na}{Na}\right)^{1/\alpha} \qquad (4)$$

In the selection experiments all the n_i and N_i are known, so α could be estimated. In the case of the inventory database both α and Na are unknown.

Nonetheless, guestimates of N_a can be used to allow an approximate estimate of the degree of bias. Intensive efforts have been made to determine the number of brook trout, lake trout, and walleye lakes in Ontario (C.H. Olver, pers. comm.) and a similar attempt is now being made for smallmouth

bass. Those estimates of N_a were used along with the values of n_a after 8867 lakes had been inventoried (=nt), to calculate approximate α's using equation (4). The results indicate the extreme degree of bias with values ranging from 13 to 45 (Table 1). Of course this rough analysis takes no account of biogeographic effects and temporal trends in the spatial allocation and intensity of sampling as the data base grew. This approach also does not account for the simultaneous action of several biases; i.e., selection for more than one fish species or other lake attributes concurrently. Difficulties involved in the detection of fish species, which may vary from species to species, are not considered here.

The presence of biases for several fish species at the same time, can lead to an overestimation of the co-occurrence of two species, as long as the proportion of the total population sampled is low; i.e., n_t small compared to N_t. This is very important if one is trying to identify lake types based on fish species combinations for assessment purposes. In the presence of bias, one may be led to consider a combination which doesn't warrant the attention. Given the high level of bias (Table 1) toward the four gamefish species, this overattention is highly likely.

The model of Manly *et al.* (1972) can be used to deal with the situation where several biases are operative at the same time. Given equations (1) to (3), assume there is a second species *b* for which the preference is β. The probability of choosing a lake with *b* present is,

$$P(b) = \frac{\beta}{1 + \beta} \qquad (5)$$

The probability of choosing a lake with both species a and b present is

$$P(a).P(b) = \frac{\alpha}{(1+\alpha)} \cdot \frac{\beta}{(1+\beta)} \qquad (6)$$

If the frequencies of occurrence of the combinations among N_t are designated as follows,

		Species b		
		Absence	Presence	Total
Species a	Absence	N_{oo}	N_{ob}	$(N_{oo}+N_{ob})$
	Presence	N_{ao}	N_{ab}	$N_{ao}+N_{ab}=Na$
	Total	$(N_{oo}+N_{ao})$	$N_{ob}+N_{ab}=Na$	N_t

the probabilities of occurrence for the four combinations after n_t lakes have been inventoried are,

$$P(oo) = \frac{(N_{oo} - n_{oo})}{D} \qquad (7)$$

$$P(ao) = \frac{\alpha\,(N_{ao} - n_{ao})}{D} \qquad (8)$$

$$P(ob) = \frac{\beta\,(N_{ob} - n_{ob})}{D} \qquad (9)$$

$$P(ab) = \frac{\alpha\beta\,(N_{ab} - n_{ab})}{D} \qquad (10)$$

where $D = (N_{oo}-n_{oo}) + \alpha(N_{ao}-n_{ao}) + \beta(N_{ob}-n_{ob}) + \alpha\beta\,(N_{ab}-n_{ab})$.

The probabilities can be integrated numerically from zero to n_t to determine expected frequencies. An example will illustrate the effect that two biases will have on co-occurrences. For instance, in the lake inventory data base about 20 percent of the lake trout lakes also contain brook trout. A hypothetical province (ratio to Ontario ≅ 1:20) would have 10000 lakes (N_T) with 100 brook trout lakes (N_a) and 100 lake trout lakes (N_b). The data base proportion of co-occurrence is assumed to be an overestimate and so I assumed there are 10 lakes (N_{ab}) containing both brook trout and lake trout. Sampling of 500 lakes, i.e., integration of probabilities from 0 to n_t=500, can be simulated with various values for the biases α and β. The results show how effective the combined biases can be in inflating estimates of the co-occurrence of the two species (Table 2). In most cases, all possible co-occurrences of a and b are included in the database after only 500 of 10000 lakes set have been inventoried.

The cumulative results of bias show a pattern of declining proportions of n_{ab}/n_t (Table 3). Thus early results will give a false impression of the frequency of occurrence of combinations. The expected percentage occurrence of both species is 0.1. When a single bias is present (α=1, β=40), the percent occurrence is 2 to 3 percent, many times expected. When two biases are operative the distortion is severe (Table 3). In the case of α=40 and β=40, nearly all the co-occurrences have been sampled after only 50 lakes have been inventoried. Of course this co-occurrence is only sustained until a large proportion of N_a and N_b have been sampled. Then n_{ab}/n_a and n_{ab}/n_b will approximate N_{ab}/N_a and N_{ab}/N_b.

ANALYSIS OF SELECTION BIAS OF THE ONTARIO LAKE INVENTORY DATABASE

Two approaches to the analysis of bias in the data base were tried: one spatial and one temporal. In the spatial analysis, the inventories were aggregated by number and lake area at the tertiary watershed level by lake area size categories. The hypothesis was that as the proportion of lakes inventoried $\frac{nt}{Nt}$ in a watershed approaches 1.0 the proportion of lakes containing a particular species $\frac{na}{nt}$ will tend toward its true value $\frac{Na}{Nt}$. The values of na/nt and nt/Nt at the tertiary watershed level were aggregated at the primary watershed level and na/nt regressed against nt/Nt. For fish species for which there is a positive bias ($\alpha > 1$), the slope should be negative; for species for which there is a negative bias ($\alpha < 1$), the slope should be positive; for species for which there is no bias ($\alpha = 1$), the slope should be zero. Results for the single-species bias model suggested that a log/log transformation might be appropriate as cumulative curves of na/nt vs nt/Nt are non-linear when $\alpha \neq 1.0$ (bias present).

The data were aggregated at the tertiary watershed level and analysed at the primary watershed level, in order to allow for regional differences in the occurrence of fish species. The results were mixed and not very useful in assessing bias. For example the linear regressions for lake trout (Table 4), had both positive and negative slopes and correlations varied considerably (0.027 to 0.610). If tertiary watersheds where na/nt was zero were excluded (this was taken as an indication that the watershed lay outside the distribution of the species in question), results didn't improve (range of r 0.034 to 0.689). One of the reasons for the poor correlations was a lack of range in the values of the independent variable (nt/Nt). Log/log regressions gave no better results, so the spatial approach to analysis was abandoned.

In the temporal analysis, the inventories were aggregated by inventory year. Cumulative tables were calculated for the occurrences of the 16 combinations of the four gamefish species. These results show much stronger evidence of the biases (Table 5) than was obtained with the spatial analyses. The percentage of inventoried lakes containing one or more of the four gamefish species has declined from over 80 to 64 percent over the period 1961-1981. Brook trout were overrepresented early on. Various combinations of gamefish peaked in the first few years. The single occurrences of walleye, smallmouth bass, and lake trout are the only categories continuing to increase as a proportion of the total.

A MODEL OF BIAS FOR THE ONTARIO LAKE INVENTORY DATABASE

The rough analysis of bias for the four gamefish species produced high bias coefficients (Table 1). At that level of bias, it is to be expected that co-occurrences of the four species will have been almost completely exhausted by now. A simple model of bias can be examined assuming that all further expected occurrences of each species will be singly rather than in mixtures (Table 6). Given these expected values of N_i for all sixteen combinations and the estimated α's from Table 1, the expected cumulative sampling pattern was computed for up to 10,000 of the 181,015 lakes in Ontario. The results (Table 7) are remarkably similar to that found with the chronological analysis of the data base (Table 5). The differences between observed (Table 5) and predicted (Table 7) suggest that non-brook trout values of α (Table 2) are overestimates and the brook trout α an underestimate. The results do suggest

that co-occurrences have been exhausted and that future lake inventories will only find single occurrences of the four gamefish species.

The basic model of bias toward four gamefish species is a reasonable description of the cumulative lake inventory database. A more thorough model fitting would require the simultaneous estimation of the N_i's and the α's though it is not clear if early data should be grouped to give larger sample size increments. Another analysis approach would be to examine the results of inventories made since 1981, to see if the trends in the percentage composition of gamefish mixtures have been sustained.

DISCUSSION

The results confirm that there are biases leading to the over-representation of certain fish species in the lake inventory data base. However, it appears that the model of Manly et al. (1972) and its extensions can be used to describe the bias. Chesson (1978) strongly endorsed the use of the Manly et al. (1972) model of selective predation, as it is one of the few that allows for differences in the frequency of occurrence of the various prey types. The results presented here strongly suggest it is an appropriate description of the biased selection of lakes for inventory.

It is not surprising that these biases should occur. These biases are in fact deliberate. Agencies charged with the management of fishery resources direct their efforts to those species which are highly prized and most exploited. However, these biases need to be understood and quantified when overall resource assessments are needed, as with the acidic deposition problem.

Given that the biases affect the inventory data base, a number of

consequences arise. Chief of these is the over-representation of co-occurrences. This causes more assessment attention to be directed their way and extrapolations down-play the role of lakes containing only one of the species. Minns (1986) found in an association analysis of the inventory database that there were three main fish assemblages - 'brook trout', 'centrarchid' and 'salmonid-percid'. Given the bias toward the top predator members of these groups, the co-occurrence of these assemblage types was likely overstated. The three types are headed by at least one of the species for which there is a strong bias. The walleye and lake trout did not appear in separate clusters as a result of the association analysis.

To extrapolate from the inventory data base, it may be necessary to treat lakes which contain more than one of the four top predators as being a complete survey and to extrapolate from the lakes which contain only one of the species.

Bias analyses of the lake inventory database are by no means complete. I have assumed that we know the number of lakes occupied by various assemblages. A more intensive analysis would involve an iterative fitting of N_i's for each assemblage and the four α's. The analyses so far do not consider how biogeographical and habitat limitations might condition the assemblage frequencies. Likely the database should be stratified to account for regional limits on the distributions of species. Legendre and Legendre (1984) performed a biogeographical analysis on the distribution of fish in Quebec, relating patterns to postglacial dispersal and topographical barriers. Such an analysis might be a useful precursor to a further analysis of bias.

As for the future conduct of lake selection for inventory in Ontario, three options are available. If the emphasis of management remains focussed on particular species, a profitable strategy might be to increase the bias even further to make an exhaustive search for the remaining lakes containing preferred species. This approach could be aided by the biogeographic analysis suggested above and by extensive surveys based on a few key parameters identified from discriminant analyses of presence/absence as presented in Minns (1986) and Beggs et al. (1985). Such parameters would likely include lake area, maximum depth, secchi depth, and elevation.

If the emphasis of management shifts further toward a more "community"-based approach with a concern for the whole freshwater resource, a stratified random method of lake selection would have to be adopted. Resource mapping already available (Cox 1978) could be exploited by having lakes selected on a watershed and size range basis. This approach would not correct for the bias in the existing lake inventory data base.

A third option would involve targetting a large number of lakes for inventory, which do not contain any of the four gamefish species. It is highly unlikely that this option would be adopted as it would mean tens of thousands of lakes being inventoried with little management application.

Whichever approach is used, it is important that the bases for lake selection be explicit and quantified where possible. As resource management becomes ever more complicated, it is important that the data collected under the auspices of inventories and surveys be suitable for extrapolation and, after correction for bias, representative of the total resource under consideration be it all lakes or all gamefish lakes.

ACKNOWLEDGEMENTS

I am pleased to recognize the continuing encouragement and support from many colleagues in the Fisheries Branch, Ontario Ministry of Natural Resources including Gail Beggs, Donna Wales, John Matusek and Jim MacLean. This work was undertaken, in part, to support efforts to develop an extensive impact assessment of acidic deposition on Ontario's freshwater fishery resources.

LITERATURE CITED

Beggs, G.L., Gunn, J.M. and Olver, C.H. 1985. The sensitivity of Ontario lake trout (Salvelinus namaycush) and lake trout lakes to acidification. Ont. Fish. Tech. Rep. Ser. No. 17, 24p.

Chesson, J. 1978. Measuring preference in selective predation. Ecology 59:211-215.

Cox, E.T. 1978. Counts and measures of Ontario lakes. Ont. Min. Natur. Resour. MS Rep. 114 p.

Dodge, D.P., Goodchild, G.A., Tilt, J.C. and Waldriff, D.G. 1981. Manual of instructions: aquatic habitat inventory surveys. 2nd edition. Ont. Min. Natur. Resour. 159 p.

Goodchild, G.A. and G.E. Gale. 1981. Aquatic habitat inventory - The Ontario approach to lake surveys. In Armantrout, N.B. (ed.). Acquisition and Utilization of Aquatic Habitat Inventory Information. Symposium held by American Fisheries Society, 376pp.

Legendre, P. and V. Legendre. 1984. Postglacial dispersal of freshwater fishes in the Quebec peninsula. Can. J. Fish. Aquat. Sci., 41:1781-1802.

Manly, B.F.J., Miller, P. and Cook, L.M. 1972. Analysis of a selective predation experiment. Amer. Nat., 106:719-736.

Minns, C.K. 1981. Acid rain: A preliminary estimate of the risk to Ontario's inland fisheries. Can. MS Rep. Fish. Aquat. Sci. 1622. 36 p.

Minns, C.K. 1984. Analysis of lake and drainage area counts and measures for selected watersheds across Ontario's shield region. Can. MS Rep. Fish. Aquat. Sci. 1748. 17 p.

Minns, C.K. 1986. Fish species assemblage structure in Ontario's lakes. Can. MS Rep. Fish. Aquat. Sci. 1898:i-vi, 1-17pp.

Ontario Ministry of Natural Resources. 1978. Designation of assessment units. Report of SPOF Working Group Number One. 66p.

Table 1. Estimates of the bias coefficient α given n_a, n_t^1, N_t^1 and 'guestimates' of N_a for four species of gamefish, using equation 4.

Fish species		Inventory sample n_a	Estimated total N_a	Relative preference[2] α
Brook trout	(B)	1764	2110	44.63
Lake trout	(L)	1634	2220	32.25
Smallmouth bass	(S)	1603	3500	14.66
Walleye	(W)	2340	6000[3]	13.01

1 - $n_t = 8867$, $N_t = 181{,}015$

2 - no preference is indicated $\alpha = 1$

3 - guestimate! 3601 indicated in 1980 walleye atlas (C.H. Olver, pers. comm.)

Table 2. Expected sample of lakes containing species a and b (n_{ab}) after sampling 500 (n_t) of 10,000 (N_t) lakes for various combinations of α and β biases using equations 7 to 10. (The line delimits those combinations where all available lakes with a and b are sampled.)

	β					
	1	5	10	20	30	40
1	0.5	2.2	3.8	6.0	7.3	8.2
5	2.2	7.0	9.0	9.9	10.0	10.0
10	3.8	9.0	9.9	10.0	10.0	10.0
20	6.0	9.9	10.0	10.0	10.0	10.0
30	7.3	10.0	10.0	10.0	10.0	10.0
40	8.2	10.0	10.0	10.0	10.0	10.0

Table 3. Percentage of sampled lakes with both species present vs. cumulative sample size with Na=100, Nb=100, Nab=10 and N_t=10,000 using equations 7 to 10.

Cumulative sample, n_t	Percentage of n_t with a and b				
	α=1 β=1	1 10	1 40	10 10	40 40
50	0.1	0.90	2.74	6.67	19.64
100	0.1	0.88	2.60	5.61	10.00
150	0.1	0.87	2.46	4.76	6.67
200	0.1	0.85	2.33	4.08	5.00
250	0.1	0.84	2.20	3.53	4.00
300	0.1	0.82	2.07	3.08	3.33
350	0.1	0.80	1.96	2.72	2.86
400	0.1	0.79	1.85	2.43	2.50
450	0.1	0.77	1.74	2.18	2.22
500	0.1	0.76	1.64	1.98	2.00

Table 4. Results of linear regressions of na/nt vs nt/Nt for lake trout by primary watershed and lake size range in Ontario with and without zero values (na/nt=0).

Lake area[1] size range	Primary[2] watershed	Tertiary watersheds	Mean nt/Nt	Mean na/nt	Regression results: Slope	Intercept	Correlation coefficient (r)
With zeroes:							
1	2	38	0.809	0.555	-0.041	0.589	-0.027
	4	39	0.671	0.137	0.258	-0.037	0.255
	5	13	0.793	0.436	-0.601	0.912	-0.443
2	2	49	0.772	0.318	0.368	0.034	0.307
	4	47	0.369	0.061	0.044	0.044	0.125
	5	17	0.432	0.377	0.652	0.095	0.538
3	2	59	0.374	0.122	-0.179	0.189	-0.337
	4	31	0.125	0.049	-0.088	0.060	-0.066
	5	12	0.058	0.231	-2.202	0.359	-0.429
4	2	56	0.105	0.011	-0.037	0.015	-0.199
	4	20	0.050	0.001	0.056	-0.002	0.610
	5	7	0.003	0.000	-	-	-
Without zeroes:							
1	2	30	0.801	0.703	0.041	0.670	0.034
	4	12	0.780	0.444	0.469	0.078	0.231
	5	11	0.755	0.515	-0.281	0.728	-0.248
2	2	41	0.790	0.380	0.469	0.010	0.340
	4	15	0.514	0.190	-0.125	0.254	-0.267
	5	14	0.471	0.458	0.483	0.230	0.451
3	2	38	0.291	0.189	-0.025	0.196	-0.038
	4	9	0.239	0.169	-0.989	0.405	-0.528
	5	9	0.052	0.308	-5.455	0.589	-0.672
4	2	10	0.045	0.061	-0.557	0.086	-0.689
	4	1	0.168	0.019	-	-	-
	5	0	-	-	-	-	-

1 - 1- 1000 ha, 2 - 100 to 999.9, 3 - 10 to 99.9, 4 - 1 to 9.9
2 - 2-Great Lakes/St. Lawrence, 4-Hudson/James Bays, 5-Nelson River

Table 5. Observed cumulative number of lake inventories, percent of lakes with one or more gamefish species and percentage of gamefish lakes with one of fifteen combinations of gamefish presence.

Year	Cumulative sample	Percent gamefish	Percent gamefish assemblage combination[1]														
			---W[2]	--S-	--SW	-B--	-B-W	-BS-	-BSW	L---	L--W	L-S-	L-SW	LB--	LB-W	LBS-	LBSW
1961	31	80.65	0.00	4.00	0.00	76.00	0.00	0.00	0.00	8.00	4.00	0.00	4.00	0.00	0.00	4.00	0.00
1962	38	84.21	3.13	3.13	9.38	68.75	0.00	0.00	0.00	6.25	3.13	0.00	3.13	0.00	0.00	3.13	0.00
1963	46	80.43	2.70	8.11	8.11	59.46	0.00	5.41	0.00	8.11	2.70	0.00	2.70	0.00	0.00	2.70	0.00
1964	66	77.27	1.96	13.73	5.88	50.98	0.00	5.88	0.00	5.88	5.88	1.96	1.96	1.96	0.00	1.96	1.96
1965	94	78.72	5.41	10.81	5.41	44.59	0.00	4.05	0.00	8.11	5.41	2.70	2.70	8.11	0.00	1.35	1.35
1966	167	81.44	4.41	5.88	4.41	51.47	0.00	2.94	0.00	8.09	5.15	4.41	2.21	8.82	0.00	1.47	0.74
1967	344	74.13	12.16	7.45	5.10	43.92	0.00	2.35	0.00	9.02	4.71	3.53	1.96	8.63	0.00	0.78	0.39
1968	915	68.09	12.36	10.59	11.56	31.94	0.48	2.09	0.00	7.38	2.89	6.58	3.21	8.67	0.16	1.77	0.32
1969	1561	66.50	14.55	10.31	10.21	32.56	0.39	1.83	0.00	7.23	3.56	6.65	3.28	7.51	0.19	1.54	0.19
1970	2256	67.77	15.63	9.35	9.48	34.60	0.26	1.96	0.07	7.98	2.88	6.08	2.88	6.87	0.13	1.50	0.33
1971	3199	65.55	16.17	10.25	8.58	34.72	0.24	2.19	0.05	8.20	2.72	5.67	2.48	6.87	0.14	1.48	0.24
1972	3985	64.77	17.32	10.58	9.45	32.16	0.23	2.25	0.04	8.76	2.79	5.66	2.67	6.39	0.12	1.39	0.19
1973	4666	64.49	18.98	11.20	9.27	30.84	0.20	2.09	0.10	8.74	2.86	5.48	2.69	5.95	0.10	1.33	0.17
1974	5264	64.49	20.29	11.08	9.10	28.92	0.18	1.91	0.12	9.54	3.15	5.68	2.74	5.83	0.12	1.18	0.15
1975	5958	64.42	20.64	11.02	9.07	27.93	0.21	1.82	0.10	10.37	3.07	5.71	2.68	5.94	0.10	1.12	0.21
1976	6434	64.94	20.90	11.30	8.90	27.12	0.22	1.68	0.10	11.37	3.09	5.77	2.63	5.58	0.10	1.05	0.22
1977	7120	64.42	22.59	11.21	8.68	26.34	0.20	1.55	0.09	11.66	2.99	5.54	2.59	5.30	0.09	1.00	0.20
1978	7677	64.24	24.88	10.87	8.41	25.24	0.22	1.50	0.08	11.80	2.96	5.25	2.47	5.09	0.08	0.95	0.18
1979	8145	64.05	26.30	10.93	8.26	24.36	0.21	1.42	0.08	11.90	3.07	5.06	2.38	4.85	0.08	0.94	0.17
1980	8550	64.04	27.07	10.81	8.07	23.85	0.20	1.35	0.07	12.29	3.12	4.88	2.34	4.79	0.07	0.91	0.16
1981	8859	64.14	27.37	10.79	7.88	23.65	0.19	1.30	0.09	12.74	3.13	4.79	2.27	4.65	0.07	0.92	0.16

1 - Combinations are specified in Table 6
2 - L-lake trout, B-brook trout, S-smallmouth bass, W-walleye

Table 6. Assumed occurrences (N_i) of all combinations of the four gamefish species in Ontario lakes.

Presence(x)/Absence(-)				n_i	+Additions = N_i	
Lake trout	Brook trout	Smallmouth bass	Walleye	in 1981		
-	-	-	-	3178	165664	168842
-	-	-	x	1555	3660	5215
-	-	x	-	613	1897	2510
-	-	x	x	448	-	448
-	x	-	-	1344	346	1690
-	x	-	x	11	-	11
-	x	x	-	74	-	74
-	x	x	x	5	-	5
x	-	-	-	725	586	1311
x	-	-	x	178	-	178
x	-	x	-	272	-	272
x	-	x	x	129	-	129
x	x	-	-	264	-	264
x	x	-	x	4	-	4
x	x	x	-	52	-	52
x	x	x	x	10	-	10
			Totals	8862	+ 172153	= 181015

Table 7. Predicted percentage of lakes with one or more of the gamefish species and percentage of lakes with one of fifteen combinations of gamefish presence/absence, based on integration of probabilities.

Cumulative sample	Percent gamefish	Gamefish assemblage combination[1]														
		----W[2]	--S-	--SW	-B--	-B-W	-SB-	-BSW	L---	L--W	L-S-	L-SW	LB--	LB-W	LBS-	LBSW
250	92.38	3.31	1.80	4.13	3.67	.30	2.28	1.34	2.06	3.56	6.12	27.98	17.14	1.52	20.46	4.33
500	89.42	4.75	2.57	5.81	5.25	.41	3.07	1.04	2.95	4.91	8.38	24.67	21.45	.89	11.62	2.24
750	86.92	6.03	3.27	7.21	6.64	.48	3.59	.76	3.74	5.91	10.03	19.24	22.97	.61	7.98	1.53
1000	84.79	7.17	3.88	8.34	7.86	.53	3.89	.59	4.43	6.61	11.13	15.16	22.63	.47	6.13	1.18
1250	82.93	8.21	4.45	9.26	8.95	.55	4.02	.48	5.05	7.07	11.81	12.44	21.35	.39	5.02	.96
1500	81.24	9.19	4.97	10.02	9.95	.56	4.03	.41	5.63	7.35	12.18	10.59	19.70	.33	4.27	.82
1750	79.68	10.12	5.48	10.64	10.88	.56	3.97	.36	6.18	7.50	12.31	9.25	18.02	.29	3.73	.72
2000	78.23	11.01	5.96	11.14	11.76	.55	3.86	.32	6.70	7.53	12.26	8.25	16.46	.26	3.32	.64
2250	76.87	11.87	6.42	11.53	12.57	.53	3.71	.29	7.18	7.47	12.07	7.46	15.08	.23	3.01	.58
2500	75.61	12.68	6.86	11.81	13.33	.51	3.55	.26	7.64	7.35	11.79	6.82	13.89	.21	2.75	.53
2750	74.43	13.47	7.28	12.02	14.03	.49	3.38	.24	8.07	7.19	11.44	6.30	12.87	.20	2.54	.49
3000	73.32	14.21	7.68	12.14	14.68	.47	3.21	.23	8.47	6.99	11.06	5.86	11.99	.18	2.36	.45
3250	72.28	14.93	8.06	12.20	15.28	.45	3.06	.21	8.85	6.78	10.66	5.49	11.23	.17	2.21	.43
3500	71.30	15.61	8.43	12.20	15.84	.43	2.91	.20	9.20	6.55	10.25	5.17	10.58	.16	2.08	.40
3750	70.37	16.26	8.77	12.16	16.34	.41	2.77	.19	9.53	6.32	9.85	4.89	10.00	.15	1.97	.38
4000	69.50	16.89	9.11	12.07	16.81	.39	2.64	.18	9.83	6.10	9.47	4.64	9.50	.14	1.87	.36
4250	68.67	17.49	9.42	11.96	17.24	.37	2.52	.17	10.12	5.88	9.10	4.42	9.05	.14	1.78	.34
4500	67.89	18.06	9.73	11.82	17.63	.36	2.41	.16	10.39	5.67	8.75	4.22	8.64	.13	1.70	.33
4750	67.14	18.61	10.02	11.66	17.98	.34	2.31	.16	10.64	5.47	8.42	4.04	8.28	.13	1.63	.31
5000	66.43	19.13	10.29	11.48	18.30	.33	2.22	.15	10.87	5.28	8.12	3.88	7.95	.12	1.57	.30
5250	65.76	19.64	10.56	11.29	18.59	.32	2.14	.14	11.09	5.10	7.83	3.74	7.65	.12	1.51	.29
5500	65.11	20.12	10.81	11.10	18.86	.31	2.07	.14	11.29	4.93	7.56	3.60	7.37	.11	1.45	.28
5750	64.49	20.59	11.06	10.90	19.10	.30	1.99	.13	11.47	4.77	7.31	3.48	7.12	.11	1.40	.27
6000	63.90	21.04	11.29	10.69	19.31	.29	1.93	.13	11.65	4.62	7.08	3.36	6.89	.10	1.36	.26
6250	63.32	21.47	11.52	10.49	19.50	.28	1.87	.13	11.81	4.48	6.86	3.26	6.67	.10	1.31	.25
6500	62.77	21.89	11.73	10.28	19.67	.27	1.81	.12	11.96	4.35	6.66	3.16	6.47	.10	1.27	.25
6750	62.24	22.29	11.94	10.08	19.83	.26	1.76	.12	12.10	4.23	6.47	3.07	6.28	.10	1.24	.24
7000	61.72	22.68	12.14	9.88	19.96	.25	1.71	.12	12.23	4.11	6.29	2.99	6.11	.09	1.20	.23
7250	61.23	23.05	12.34	9.68	20.08	.25	1.67	.11	12.35	4.01	6.13	2.91	5.95	.09	1.17	.23
7500	60.74	23.42	12.53	9.49	20.18	.24	1.62	.11	12.46	3.90	5.97	2.83	5.80	.09	1.14	.22
7750	60.27	23.77	12.71	9.30	20.27	.24	1.58	.11	12.56	3.81	5.82	2.76	5.65	.09	1.11	.21
8000	59.81	24.12	12.88	9.12	20.34	.23	1.55	.10	12.66	3.72	5.68	2.70	5.52	.08	1.09	.21
8250	59.37	24.45	13.05	8.95	20.41	.22	1.51	.10	12.75	3.63	5.55	2.63	5.39	.08	1.06	.20
8500	58.93	24.78	13.22	8.78	20.46	.22	1.48	.10	12.83	3.55	5.43	2.58	5.27	.08	1.04	.20
8750	58.51	25.09	13.38	8.61	20.50	.21	1.45	.10	12.90	3.48	5.31	2.52	5.16	.08	1.02	.20
9000	58.09	25.40	13.53	8.45	20.53	.21	1.42	.10	12.97	3.40	5.20	2.47	5.05	.08	.99	.19
9250	57.69	25.70	13.69	8.30	20.55	.21	1.39	.09	13.04	3.34	5.10	2.42	4.95	.07	.97	.19
9500	57.29	26.00	13.83	8.15	20.57	.20	1.36	.09	13.10	3.27	5.00	2.37	4.85	.07	.96	.18
9750	56.90	26.28	13.98	8.01	20.57	.20	1.33	.09	13.15	3.21	4.90	2.33	4.76	.07	.94	.18
10000	56.51	26.56	14.12	7.87	20.57	.19	1.31	.09	13.20	3.15	4.81	2.28	4.67	.07	.92	.18

1 - Combinations are specified in Table 6.
2 - See Table 5